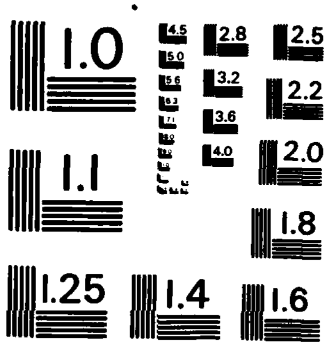
AD-A160 967 COMPARISONS OF LEAST SQUARES AND ERRORS-IN-VARIABLES REGRESSION WITH SPEC..(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER R J CARROLL ET AL. SEP 85 1/1

UNCLASSIFIED MRC-TSR-2866 DAAG29-80-C-0041 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

2

MRC Technical Summary Report #2866

COMPARISONS OF LEAST SQUARES AND ERRORS-IN-VARIABLES REGRESSION, WITH SPECIAL REFERENCE TO RANDOMIZED ANALYSIS OF COVARIANCE

Raymond J. Carroll, Paul Gallo and
Leon Jay Gleser

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

September 1985

(Received November 27, 1984)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

Air Force Office of
Scientific Research
Washington, DC 20332

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

COMPARISONS OF LEAST SQUARES AND ERRORS-IN-VARIABLES REGRESSION, WITH SPECIAL REFERENCE TO RANDOMIZED ANALYSIS OF COVARIANCE

Raymond J. Carroll[1,*], Paul Gallo[2,**] and Leon Jay Gleser[3]

Technical Summary Report #2866
September 1985

ABSTRACT

In an errors-in-variables regression model, the least squares estimate is generally inconsistent for the complete regression parameter but can be consistent for certain linear combinations of this parameter. The authors ~~We~~ explore the conjecture that, when least squares is consistent for a linear combination of the regression parameter, it will be preferred to an errors-in-variables estimate, at least asymptotically. The conjecture is false, in general, but it is true for important classes of problems. One such problem is a randomized two-group analysis of covariance, upon which this document ~~we~~ focuses.

Keywords: maximum likelihood estimation.

AMS (MOS) Subject Classifications: Primary 62J05; Secondary 62F10

Key Works: regression, measurement error, randomized studies, functional models, structural models, analysis of covariance, asymptotic theory

Work Unit Number 4 (Statistics and Probability)

---

[1]Raymond J. Carroll, Professor of Statistics, University of North Carolina, Chapel Hill.
[2]Paul Gallo, Preclinical Group Leader, Lederle Laboratories.
[3]Leon Jay Gleser, Professor of Statistics, Purdue University.
*Supported by Air Force Office of Scientific Research Contract AFOSR-F-49620-82-C-0009 and sponsored by the United States Army under Contract No. DAAG29-80-C-0041.
**Supported by the National Science Foundation.

SIGNIFICANCE AND EXPLANATION

In an errors-in-variables regression model, the least squares estimate is generally inconsistent for the complete regression parameter but can be consistent for certain linear combinations of this parameter. We explore the conjecture that, when least squares is consistent for a linear combination of the regression parameter, it will be preferred to an errors-in-variables estimate, at least asymptotically. The conjecture is false, in general, but it is true for important classes of problems. One such problem is a randomized two-group analysis of covariance, upon which we focus.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

COMPARISONS OF LEAST SQUARES AND ERRORS-IN-VARIABLES REGRESSION, WITH SPECIAL REFERENCE TO RANDOMIZED ANALYSIS OF COVARIANCE

Raymond J.Carroll[1,*], Paul Gallo[2,**] and Leon Jay Gleser[3]

1. Introduction

The literature on the problem of linear regression when some of the predictors are measured with error is substantial, see for example, Reilly and Patino-Leal (1981). Recent work includes the theoretical study of Gleser (1981) and the important practical shrinkage suggestions of Fuller (1980). See also Anderson (1984) and Healy (1980).

A subarea of this literature concerns two-group analysis of covariance when some of the predictors are measured with error, see for example Lord (1960), Cochran (1968), DeGracie and Fuller (1972) and Cronbach (1976).

Lord (1960) discusses the case of one covariate measured with error. He notes that it may "happen ... that the usual covariance analysis (least squares) will fail to detect a statistically significant difference between groups ... when such a difference actually exists and can be detected by proper statistical procedures." He also gives a numerical *example of this phenomenon*.

Cochran (1968) and DeGracie and Fuller (1972) discuss two group analysis of covariance, providing in particular some discussion of the case that the true values of the covariates are themselves random variables; this is usually called a "structural" model in the literature. They show that if the covariables are unbalanced as might happen in an observational study, then the measurement error will cause least squares to inconsistently

---

[1]Raymond J. Carroll, Professor of Statistics, University of North Carolina, Chapel Hill.
[2]Paul Gallo, Preclinical Group Leader, Lederle Laboratories.
[3]Leon Jay Gleser, Professor of Statistics, Purdue University.
*Supported by Air Force Office of Scientific Research Contract AFOSR-F-49620-82-C-0009 and sponsored by the United States Army under Contract No. DAAG29-80-C-0041.
**Supported by the National Science Foundation.

estimate the true treatment difference. In the sense of asymptotics, when the covariables are unbalanced one should then correct for measurement error if it is substantial; a global small sample statement of this type cannot be made.

Now consider a completely randomized study, where the covariables will be balanced on average across the two treatments. In this case, Cochran (1980) and DeGracie and Fuller (1980) indicate that least squares will consistently estimate the treatment difference. The question which remains to be answered is "Should we correct for measurement error when the least squares estimate consistently estimates the treatment effect?" It is the purpose of this note to partially answer this question. Using large sample distribution theory, we show that in a balanced, completely randomized study with measurement error in the covariables, the least squares estimate of the treatment difference will be generally preferred when compared to a particular errors-in-variables regression estimator. It turns out that this result can be generalized, so that in a large class of problems, when least squares is consistent for a linear combination of the regression parameter, it will be preferred, at least asymptotically. Further, for a smaller but not insubstantial class of problems, when least squares is consistent for a linear combination of the regression parameter, it is the maximum likelihood estimate of this linear combination, taking the consistency into account.

2. The Normal Case with no Replication: Technical Background

A special case of considerable interest occurs when all errors are normally distributed and no replicates of the variables measured with error are available. The general model considered here, which includes the analysis of covariance as a special case, is given by

$$\begin{aligned} Y &= X_1\beta_1 + X_2\beta_2 + \varepsilon \ , \\ C &= X_2 + U \\ \beta &= [\beta_1^T, \beta_2^T]^T \ . \end{aligned} \tag{2.1}$$

Here, Y and ε are $(N \times 1)$ vectors, X_1 is an $(N \times p)$ matrix observed without error and X_2 is an $(N \times q)$ matrix of true values which we cannot observe exactly. Rather, we observe C. The rows of the matrix (U,ε) will be assumed to be jointly normally distributed with mean zero and unknown covariance Σ .

In comparing least squares and errors-in-variables methods, we must pick a representative member of the latter class. In the main, we will do this by following Gleser (1981) for the case that no replicated estimates of X_2 are available; the replicated case will be discussed at the end of the article. Gleser studies the functional model in which X_1, X_2 are considered as fixed constants. A special case of his model assumes that there is a known matrix Σ_0 and an unknown constant σ^2 for which

$$\Sigma = \sigma^2\Sigma_0 = \sigma^2 \begin{pmatrix} \Sigma_{uo} & 0 \\ 0 & 1 \end{pmatrix} . \tag{2.2}$$

If Σ_u is the covariance matrix of the rows of U, then in (2.2) we are assuming that we know the ratio of the elements of Σ_u to σ_ε^2 , the variance of the elements of ε . Gallo (1982) exhibits the maximum likelihood estimate of β, which is given in Appendix 1.

He also proves the following:

<u>Theorem 1</u> (From Gallo (1982)). Suppose that

$$\Delta = \lim_{N\to\infty} N^{-1}(X_1,X_2)^T(X_1,X_2) \tag{2.3}$$

exists and is positive definite. Then if $\hat{\beta}_M$ is the functional maximum likelihood estimate, $N^{1/2}(\hat{\beta}_M-\beta)$ is asymptotically normally distributed with zero mean and covariance

$$\mathrm{Cov}(\hat{\beta}_M) = d\left\{\Delta^{-1} + \Delta^{-1}\begin{pmatrix} 0 & 0 \\ 0 & \Omega \end{pmatrix}\Delta^{-1}\right\}, \quad \text{where}$$

$$d = [\beta_2^T,-1]\ddagger[\beta_2^T,-1]^T$$

$$\Omega^{-1} = [I,\beta_2]\ddagger^{-1}[I,\beta_2]^T .$$

3. Analysis of Covariance

Consider a completely randomized two group analysis of covariance, with covariables subject to error. Formally, this problem can be subsumed into the more general structure (2.1) by letting X_2 be the covariables and

$$X_1^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ s_1 & s_2 & \cdots & s_N \end{pmatrix} , \quad \beta_1 = \begin{pmatrix} \mu \\ \alpha \end{pmatrix} , \quad X_2^T = (x_{21} x_{22} \cdots x_{2N}) . \tag{3.1}$$

We will let the s_i represent treatment assignment, standardized to have mean zero and variance one. Specifically,

$$\begin{aligned} s_i &= -\{(1-\pi)/\pi\}^{1/2} \quad \text{with probability } \pi , \\ &= \{\pi/(1-\pi)\}^{1/2} \quad \text{with probability } (1-\pi) , \end{aligned}$$

where π is the probability of assignment into treatment #1. The treatment difference is then $\alpha/\{(1-\pi)\pi\}^{1/2}$. We shall treat the true covariables as if they were random variables independent of treatment assignment and with covariance matrix Σ_x. In order to facilitate discussion we do not write down detailed assumptions; rather, we will apply Theorem 1 formally, while we will assume appropriate conditions to compute the limiting distribution of least squares. A more general result is given in Section 5.

The following result shows that as long as treatment assignment is random, asymptotically least squares is the better estimate of the treatment effect α, because both estimates are asymptotically normal with the same mean and least squares has the smaller variance.

Theorem 2 The least squares estimate $\hat{\alpha}_L$ is asymptotically normally distributed with mean α and variance $\sigma^2(L)/N$, where

$$\sigma^2(L) = \sigma^2 + \beta_2^T \Sigma_u \beta_2 - \beta_2^T \Sigma_u (\Sigma_x + \Sigma_u)^{-1} \Sigma_u \beta_2 . \tag{3.2}$$

The functional estimate $\hat{\alpha}_M$ has the same asymptotic mean but has asymptotic variance $\sigma^2(M)/N$, where

$$\sigma^2(M) = \sigma^2 + \beta_2^T \Sigma_u \beta_2 . \tag{3.3}$$

It is reasonable to conjecture that complete randomization is not necessary for Theorem 2. For example, one might randomize in blocks or use alternative balancing schemes, see Wei (1978). This conjecture is worth further study, and might be facilitated by use of equation (A.7) in the appendix.

It should be noted that in a balanced randomized study, the usual t-test for treatment effect has correct nominal level asymptotically. Thus, from both an estimation and inferential standpoint, for large samples least squares will be preferred over the functional estimate.

The folklore of the area indicates that, asymptotically, least squares estimates are biased but generally less variable than errors-in-variables estimates. The situation that has been considered in this section is one in which the least squares estimate of treatment effect has no asymptotic bias, so that it was reasonable to conjecture a preference for least squares. We shall show in Section 5, however, that it is not true that consistency of least squares for a linear combination of β always means asymptotic preferability of least squares, although it is true for a large class of problems.

4. Some Extensions

In some instances an assumption such as (2.2) will not be tenable so that a functional estimate cannot be computed. There are many ways out of this dilemma. One is to take independent replicates C_1, C_2 of X_2 in (2.1). One can compute the normal theory functional estimate in this case and obtain a result similar to Theorem 2, but more general in the sense that the underlying random variables need not actually be normally distributed. The computation of this functional estimate and its asymptotic distribution theory are available in, for example, Gallo (1982).

There are instances other than randomized two-group analysis of covariance in which certain linear combinations of the least squares estimate are consistent for the same linear combinations of the parameter. Consider the model (2.1) with $\beta^T = (\beta_1^T, \beta_2^T)$ in which it is desired to estimate the parameter $\gamma^T\beta$, where $\gamma^T = (\gamma_1^T, \gamma_2^T)$. Partitioning Δ in (2.3) into components Δ_{ij}, informally the least squares estimate satisfies

$$\hat{\beta}_L = \left((X_1,C)^T(X_1,C)\right)^{-1}(X_1,C)^T Y$$

$$\xrightarrow{P} \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} + \Sigma_u \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ -\Sigma_u \beta_2 \end{pmatrix} + \beta \; . \tag{4.1}$$

This leads us to a result which is proved formally by Gallo (1982):

<u>Theorem 3</u> The least squares estimate $\gamma^T\hat{\beta}_L$ is consistent for $\gamma^T\beta$, i.e., converges in probability to $\gamma^T\beta$ for all $\beta, \sigma^2, \Sigma_u$, if and only if

$$\gamma_2^T = \gamma_1^T \Delta_{11}^{-1} \Delta_{12} \; . \tag{4.2}$$

To see the relevance of Theorem 3, consider once again the two group analysis of covariance of Section 3. Here we have

$$\gamma_2 = 0, \; \gamma_1^T = (0,1), \; \Delta_{11} = \text{Identity},$$

$$\Delta_{12}^T = (\text{plim } N^{-1} e_*^T X_2, \; \text{plim } N^{-1} s_*^T X_2) \; ,$$

where

$$s_*^T = (s_1\ s_2 \ldots s_N),\ e_*^T = (1\ 1 \ldots 1)\ .$$

Theorem 3 says that the least squares estimate of the parameter α will be consistent for α only when

$$N^{-1} s_*^T X_2 \xrightarrow{P} 0\ . \tag{4.3}$$

Note that (4.3) is simply the requirement that the covariables be mean balanced across the two treatments. Theorem 3 indicates that only when we have such balance will the least squares treatment effect estimate be consistent.

5. Further Comparisons of Least Squares and Maximum Likelihood

On the basis of the previous discussion one might reasonably conjecture that when least squares is consistent for $\gamma^T\beta$ then asymptotically it must be better than the functional estimate of $\gamma^T\beta$. In model (2.1), our special cases such as analysis of covariance have relied upon a degree of orthogonality between X_2 and the non-intercept components of X_1. Specifically, for least squares one must deal with the following,

$$N^{1/2}\left(\gamma_1^T(X_1^TX_1)^{-1}X_1^TX_2-\gamma_2^T\right) , \tag{5.1}$$

which is a term in the linear expansion, see (A.2). For example, suppose X_1 and X_2 are very strongly orthogonal in the sense that

$$N^{1/2}\left\{\gamma_1^T(X_1^TX_1)^{-1}X_1^TX_2\right\} \longrightarrow 0 . \tag{5.2}$$

Since X_2 is unknown (5.2) can never be verified and in fact fails in a randomized analysis of covariance with $\gamma_2 = 0$, $\gamma_1 = (0\ 1)$. However, (5.2) does imply (4.2) and consistency of least squares. It is fairly easy to show that if (5.2) holds, then the least squares estimate of $\gamma^T\beta$ can be no worse than the functional estimate, at least asymptotically.

Further investigation of the conjecture is rather technical. The conjecture is false for the functional case, in general. Consider an analysis of covariance in which the treatment assignments $\{s_i\}$ occurs in the fixed sequence $\{-1,+1,-1,+1,\ldots\}$. Let the covariables $\{x_i\}$ be fixed. In a variety of circumstances, it can be shown tht the least squares estimate $\hat{\alpha}_L$ of the treatment effect α in model (3.1) satisfies

$$N^{1/2}(\hat{\alpha}_L-\alpha) \doteq A_1V + A_2N^{-1/2}\sum_{i=1}^{N} s_ix_i , \tag{5.3}$$

where A_1 and A_2 are constants and V is a weighted sum of independent observations not depending on $\{s_ix_i\}$. Equation (5.3) shows that asymptotic normality with mean zero of the least squares estimate when centered at the treatment effect α requires that

$$N^{-1/2}\sum_{i=1}^{N} s_ix_i \tag{5.4}$$

either converge in probability to zero or that (5.4) be itself asymptotically normally distributed. For the functional model, the latter case is not possible while the former case is (5.2). Since (5.4) can diverge as $N \to \infty$ with (4.2) still holding, for the functional case this means that least squares will not be always better asymptotically than maximum likelihood when least squares is consistent.

Now consider the structural case in which the rows of matrix (X_1, X_2) are independent and identically distributed. The first column of X_1 is a column of ones and X_1 is observed exactly, while X_2 is observed with error as in model (2.1). Suppose we are interested in estimating a linear combination $\gamma^T\beta$ for which least squares is known to be consistent, i.e., (4.2) holds.

Theorem 4 Make the following assumption:

Given X_1, the rows of $R = X_2 - X_1\Delta_{11}^{-1}\Delta_{12}$ are independent and identically distributed with mean zero and covariance $\quad$ (5.5)

$$\Delta_{22.1} = \Delta_{22} - \Delta_{21}\Delta_{11}^{-1}\Delta_{12} .$$

Further, suppose that R is distributed independently of ε and U. If we define

$$\Lambda = (\Delta_{22.1} + \Sigma_u)^{-1} ,$$

we have that the least squares and functional maximum likelihood estimates are asymptotically normally distributed with mean zero and variances $\sigma^2(L)/N$, $\sigma^2(M)/N$ respectively, where $\sigma^2(L) < \sigma^2(M)$. In fact

$$\sigma^2(L) = \sigma^2(M) - (\gamma_1{}^T\Delta_{11}^{-1}\gamma_1)\beta_2{}^T\Sigma_u\Lambda\Sigma_u\beta_2$$

$$\sigma^2(M) = (\gamma_1{}^T\Delta_{11}^{-1}\gamma_1)(\sigma^2 + \beta_2^T\Sigma_u\beta_2) . \qquad \square$$

The proof of Theorem 4 is given in the Appendix 2. Note that it includes Theorem 2 as a special case because when X_1 is distributed independently of X_2, then (5.5) holds. That Theorem 4 may not hold when assumption (5.5) is violated is sketched in Appendix 3.

It may be considered a bit unfair to compare least squares to a "maximum likelihood estimator" which does not take into account the consistency of least squares. It turns

out that, under normality assumptions, the maximum likelihood estimate of $\gamma^T\beta$ when it is known that least squares is consistent for $\gamma^T\beta$ is simply the least squares estimate of $\gamma^T\beta$. Specifically, we have the following.

<u>Theorem 5</u> *Suppose that, given* X_1, *(5.5) holds and the rows of* $R = X_2 - X_1\Delta_{11}^{-1}\Delta_{12}$ *are normally distributed independently of* ε *and* U. *Then the maximum likelihood estimate of* $\gamma^T\beta$ *given* X_1 *and subject to (4.2) is simply the least squares estimate of* $\gamma^T\beta$.

6. Conclusion

In a particular errors-in-variables regression model, we have shown that least squares will often be asymptotically more efficient than a particular functional regression estimate, when the former is known to be consistent. This happens in particular when those variables X_2 subject to error are distributed independently of those variables X_1 measured without error, or more generally when X_2 follows a linear regression in X_1. An important special case of this least squares preference phenomenon is a randomized analysis of covariance where one wants to estimate the treatment effect. Finally, if the linear regression of X_2 on X_1 follows a multinormal distribution, and if it is known that least squares is consistent for the linear combination $\gamma^T\beta$, then least squares is the maximum likelihood estimate for $\gamma^T\beta$.

Acknowledgement

The authors thank Robert D. Abbott and Leonard A. Stefanski for helpful discussions.

Appendix 1: The maximum likelihood estimator for model (2.1)

Define

$$L = I - X_1(X_1^T X_1)^{-1} X_1^T$$

$$W = [C\ Y]^T L [C\ Y] \ .$$

Let θ be the smallest eigenvalue of $\Sigma_0^{-1} W$, where Σ_0 is given in (2.2).

Define

$$C_* = [X_1\ C] = [X_1\ X_2 + U] ,$$

$$D = C_*^T C_* - \theta \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{uo} \end{pmatrix} .$$

The matrix D is non-singular with probability one, and the functional estimate is

$$\hat{\beta}_M = D^{-1} C_*^T Y \ .$$

The calculation of $\hat{\beta}_M$ is derived by Gallo (1982) and relies on similar work of Gleser (1981) and Healy (1980).

Appendix 2: The asymptotic distribution of least squares

The following general result can be justified formally and is at the heart of the analysis of covariance calculations. We sketch herein a proof without stating all the necessary regularity conditions. Recall $e_*^T = (1\ 1\ \ldots\ 1)$.

Lemma A Define

$$\Delta_{22.1} = \Delta_{22} - \Delta_{21}\Delta_{11}^{-1}\Delta_{12}\ , \qquad \Lambda = (\Delta_{22.1} + \Sigma_u)^{-1}\ ,$$

and suppose that γ satisfies (4.2) as well as

$$N^{-1/2} X_1{}^T(R + U) = O_p(1)\ , \tag{A.1}$$

where $R = X_2 - X_1\Delta_{11}^{-1}\Delta_{12}$. Then the least squares estimate satisfies

$$\begin{aligned} &N^{1/2}\gamma^T(\hat{\beta}_L - \beta) \\ &= N^{-1/2}\gamma_1{}^T\Delta_{11}^{-1}X_1{}^T(\epsilon - U\,\beta_2) \\ &+ N^{-1/2}\gamma_1{}^T\Delta_{11}^{-1}X_1{}^T\ (R + U)\xi + O_p(1)\ , \end{aligned} \tag{A.2}$$

where

$$\xi = \Lambda\Sigma_u\beta_2\ .$$

Proof (Sketch): Define $C_* = [X_1\ X_2 + U]$. Then

$$\frac{(C_*^T C_*)}{N}(\hat{\beta}_L - \beta) + \begin{pmatrix} 0 \\ \Sigma_u\beta_2 \end{pmatrix} = C_*^T(\epsilon - U\beta_2)/N + \begin{pmatrix} 0 \\ \Sigma_u\beta_2 \end{pmatrix} \tag{A.3}$$

Multiply both sides of (A.3) by $N^{1/2}\gamma^T(C_*^T C_*/N)^{-1}$ to get

$$\begin{aligned} N^{1/2}\gamma^T(\hat{\beta}_L - \beta) &= N^{1/2}\gamma^T(C_*^T C_*/N)^{-1}\left(C_*^T(\epsilon - U\beta_2)/N + \begin{pmatrix} 0 \\ \Sigma_u\beta_2 \end{pmatrix}\right) \\ &- N^{1/2}\gamma^T(C_*^T C_*/N)^{-1}\begin{pmatrix} 0 \\ \Sigma_2\beta_2 \end{pmatrix}\ . \end{aligned} \tag{A.4}$$

by Slutsky's Theorem the first term on the right hand side of (A.4) equals

$$N^{1/2}\gamma^T\left(\Delta + \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_u \end{pmatrix}\right)^{-1}\left(C_*^T(\varepsilon - U\beta_2)/N + \begin{pmatrix} 0 \\ \Sigma_u\beta_2 \end{pmatrix}\right) + o_p(1) \tag{A.5}$$

$$= N^{-1/2}\gamma_1^T\Delta_{11}^{-1}X_1^T(\varepsilon - U\beta_2) + o_p(1)\ ,$$

which is the same as the first term on the right hand side of (A.2). The second term in (A.4) is

$$N^{1/2}\gamma_1^T(X_1^TX_1)^{-1}X_1^T(R + U)W\,\Sigma_u\beta_2\ , \tag{A.6}$$

where

$$W \xrightarrow{P} (\Delta_{22.1} + \Sigma_u)^{-1}\ , \qquad (X_1^TX_1/N) \longrightarrow \Delta_{11}\ .$$

By (A.1), this completes the proof.

One should note that (A.1) is satisfied in the randomized two group analysis of covariance of Section 3.

Using Lemma A and writing for the analysis of covariance

$$X_2^T = (x_{21}\ x_{22}\ \cdots\ x_{2N})\ ,$$

$$U^T = (u_1\ u_2\ \cdots\ u_N)\ ,$$

$$m_2 = N^{-1}\sum_{i=1}^{N} x_{2i}$$

we see that for the discussion in Section 3,

$$N^{1/2}(\hat{\alpha}_L - \alpha) = N^{-1/2}\sum_{i=1}^{N} s_i\{\varepsilon_i + (\eta - \beta_2)^TU_i + \eta^T(x_{2i} - m_2)\}\ , \tag{A.7}$$

$$\eta = (\Lambda + \Sigma_u)^{-1}\Sigma_u\beta_2 = (\Sigma_x + \Sigma_u)^{-1}\Sigma_u\beta_2\ .$$

The expression (A.7) shows why Theorem 2 may apply to alternative randomization schemes.

Proof of Theorem 4 The form of $\sigma^2(M)$ follows directly from Theorem 1. The form of $\sigma^2(L)$ follows from (A.2) and the assumptions of the Theorem.

Proof of Theorem 5

First assume that $\Delta_{11}^{-1}\Delta_{12}$ is known. Define

$$\pi = (I,\ \Delta_{11}^{-1}\Delta_{12})\beta$$

$$\ell^2 = \sigma^2 + \beta_2^T\ddagger_u\beta_2 - \beta_2^T\ddagger_u\Lambda\ddagger_u\beta_2 \ .$$

Given (X_1,C), we have

$$(Y|X_1,C) = X_1\pi + S\Lambda\Delta_{22.1}\beta_2 + F \ , \tag{A.8}$$

where $S = C - X_1\Delta_{11}^{-1}\Delta_{12}$ and the rows of F are independent normal random variables with mean zero and variances ℓ^2.

If we define

$$\ddagger_x = \sigma^{-2}\Delta_{22.1}$$

$$\xi = \Lambda\Delta_{22.1}\beta_2$$

$$L = \Lambda^{-1}$$

then the mapping of β_1, β_2, σ^2, $\Delta_{22.1}$ to π, ξ, σ^2, L is one-to-one from the space $\{\sigma^2 > 0,\ \Delta_{22.1} > 0\}$ to the space $\{\sigma^2 > 0,\ L - \sigma^2\ddagger_{uo} > 0\}$. One next shows that the map π, ξ, L, σ^2 to π, ξ, L, ℓ^2 is also one-to-one onto the space $\{\ell^2 > 0,\ L > 0\}$.
However, the maximum likelihood estimates of π and ξ are seen from (A.8) to be

$$\{(X_1,S)^T\ (X_1,S)\}^{-1}\ (X_1,S)^TY \ .$$

Since the column space of (X_1,C) is the same as the column space of (X_1,S), it follows that, given $(X_1,\ S,\ \Delta_{11}^{-1}\Delta_{12})$, the maximum likelihood and least squares estimates of π coincide, i.e.,

$$\hat{\pi}(\text{mle}) = (I,\Delta_{11}^{-1}\Delta_{12})\hat{\beta}_L \ .$$

This means that $\gamma^T\hat{\beta}_L$ is the maximum likelihood estimate of $\gamma^T\pi$, given X_1,S and $\Delta_{11}^{-1}\Delta_{12}$. Since, under (4.2), $\gamma^T\pi = \gamma^T\beta$, the proof is complete.

Appendix 3. A Counterexample

If the rows of (X_1,X_2) are independent and identically distributed but (5.5) does not hold, it is possible to construct a counterexample to Theorem 4. The way to do this is to consider model (3.1) but with the pairs $\{(s_i,x_i)\}$ satisfying

$$Es = Es^3 = 0 \ , \qquad Es^2 = Ex^2 = 1 \ ,$$

$$Ex = \theta_1 \neq 0 \ , \qquad Esx = \theta_2 \neq 0 \ ,$$

$$\gamma_1^T = (-\theta_2,\theta_1) \ , \qquad \gamma_2 = 0 \ .$$

In this case, the expansion (A.2) still holds and the last term in this expansion is

$$N^{1/2} \sum_{i=1}^{N} (\theta_1 s_i - \theta_2) \begin{pmatrix} x_i - \theta_1 - \theta_2 s_i \\ + U_i \end{pmatrix} \xi \ . \tag{A.9}$$

The key to Theorem 4 is that, under (5.5), $(x_i - \theta_1 - \theta_2 s_i + U_i)$ has mean zero and variance Λ^{-1}. Without assumption (5.5), one can see that while (A.9) has mean zero, its variance can depend on the fourth moment of $\{s_i\}$. By manipulating this fourth moment appropriately, Theorem 4 can be made to fail.

REFERENCES

Anderson, T. W. (1984). Estimating linear statistical relationships. Ann. Statist. 12, 1-45.

Cochran, W. G. (1968). Errors of measurement in statistics. Technometrics 10, 637-666.

Cronbach, L. J. (1976). On the design of educational measures. In Advancement in Psychological and Educational Measurement, ed. D. N. M. DeGruijter and L. J. Th. van der Kamp, John Wiley and Sons, New York.

Debracie, J. S. and Fuller, W. A. (1980). Estimation of the Slope and Analysis of Covariance when the Concomitant Variable is Measures with Error. J. Amer. Statist. Assoc. 67, 930-937.

Lord, F. M. (1960). Large-sample covariance analysis when the control variable is fallible. J. Am. Statist. Assoc. 55, 307-321.

Fuller, W. A. (1980). Properties of some estimators for the errors-in-variables model. Ann. Statist. 8, 407-422.

Gallo, P. P. (1982). Properties of estimators in errors-in-variables regression models. Ph.D. dissertation, University of North Carolina at Chapel Hill.

Gleser, L. J. (1981). Estimation in a multivariate "errors in variables" regression models: large sample results. Ann. Statist. 9, 24-44.

Healy, J. D. (1980). Maximum likelihood estimation of a multivariate linear functional relationship. J. Multivariate Anal. 10, 243-251.

Reilly, P. M. and Patino-Leal, H. (1981). A Bayesian study of the errors-in-variables model. Technometrics 23, 221-231.

Wei, L. J. (1978). Adaptive biased coin designs. Ann. Statist. 6, 92-100.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER #2866	2. GOVT ACCESSION NO. AD-A160967
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) COMPARISONS OF LEAST SQUARES AND ERRORS-IN-VARIABLES REGRESSION, WITH SPECIAL REFERENCE TO RANDOMIZED ANALYSIS OF COVARIANCE	5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Raymond J. Carroll, Paul Gallo and Leon Jay Gleser	8. CONTRACT OR GRANT NUMBER(s) AFOSR-F-49620-82-C-0009 DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS (See Item 18 below)	12. REPORT DATE September 1985
	13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

Air Force Office of
Scientific Research
Washington, DC 20332

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

regression, measurement error, randomized studies, functional models, structural models, analysis of covariance, asymptotic theory

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In an errors-in-variables regression model, the least squares estimate is generally inconsistent for the complete regression parameter but can be consistent for certain linear combinations of this parameter. We explore the conjecture that, when least squares is consistent for a linear combination of the regression parameter, it will be preferred to an errors-in-variables estimate, at least asymptotically. The conjecture is false, in general, but it is true for important classes of problems. One such problem is a randomized two-group analysis of covariance, upon which we focus.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

12-85

DTIC